



## Webニュースを用いた未来情報年表の自動構築

|      |   |
|------|---|
| 著者   | 吉田 光男, 乾 孝司, 山本 幹雄  |
| 内容記述 | 第3回楽天研究開発シンポジウムポスターセッション, 楽天タワー2号館, 2010年12月18日                                   |
| 発行年  | 2010-12   |
| URL  | <a href="http://hdl.handle.net/2241/112996">http://hdl.handle.net/2241/112996</a> |

# Webニュースを用いた未来情報年表の自動構築

吉田光男, 乾孝司, 山本幹雄 (筑波大学)

## 未来予測の基盤システムを目指します

### 背景と目的

- 未来予測は人々の普遍的欲求
- 未来情報を含む大量のテキスト
  - 未来情報データベースを自動構築
- 情報検索システムでキーワード選定が難しい
  - 年表形式による提示

### 取り組んでいく課題

- 未来予測の基盤システム
  - 特定の情報にフォーカスしない
  - 様々な角度からの情報提供
  - 信頼性の担保
- 情報提示方法の改善
  - キーワード選定などの負担軽減
  - 情報の要約

## 提案手法の概要

### ニュース収集

Ceek.jp News

### 対象記事抽出

西暦4桁に着目

### 未来情報抽出

### 要約

#### 未来情報は従属的である場合が多い

この夏の猛暑の影響で、夏季終了時点での北極の海水面積が観測史上3番目の小ささとなったとの研究報告が15日、相次いで発表された。地球温暖化によって2030年9月には北極海水が消滅する恐れさえあるとしている。  
(AFPBB News『今夏の北極海水面積、史上3番目の小ささ 米研究』リード文)

#### 係り受けを調べて要約

- 未来を示す「YYYY年」を含む文節を探す
- (1)から係る動詞を含む最初の文節を探す
- 動詞にかかる文節を再帰的に探す
- (2)(3)を文節順に出力する

動詞は基本形に

#### 未来情報年表

| 西暦    | イベント                    |
|-------|-------------------------|
| 2030年 | 北極海水が消滅する               |
| 2030年 | 温室効果ガスの排出権を1990年比で40%削減 |
| 2050年 | 中国の老齢人口が30%を超える         |

#### 文の中で現在と未来の事象が混在

中国の民間機構・新探健康発展研究センターの呉宜群副主任は、このほど「タバコの流行がこのまま進めば、中国では、タバコが原因で死亡する人の数は2050年までに年間300万人になり、合わせて1億人に上る恐れがある」と明らかにしました。  
(中国放送局『中国、2050年までにタバコによる死者1億人の恐れ』抜粋)

#### 係り受け解析の例

呉宜群副主任は、このほど「タバコの流行がこのまま進めば、中国では、タバコが原因で死亡する人の数は2050年までに年間300万人になり、合わせて1億人に上る恐れがある」と明らかにしました。

## 構築例・課題

- 2020年 世界の自動車市場で15%のシェアを確保する
- 2020年 温室効果ガスを1990年比で25%削減する
- 2020年 [2020年までの]核廃絶を訴える
- 2020年 [2020年夏季五輪の]誘致を目指す
- 2029年 次回の接近は[2029年に]なる
- 2030年 地球温暖化によって[2030年9月には]北極海水が消滅する恐れさえある

### 要約が適切か?

「〇〇は」を残す

- 主題が欠ける傾向
- 時間情報が連体修飾語場合の変形

### 情報の信頼性

数年前と現在の予測

- 蓄積による信頼性の担保
- 映画などのフィクションの除去

## futuretimeline.jp

ウェブサービスとして公開  
フィードバックをもとに改良

